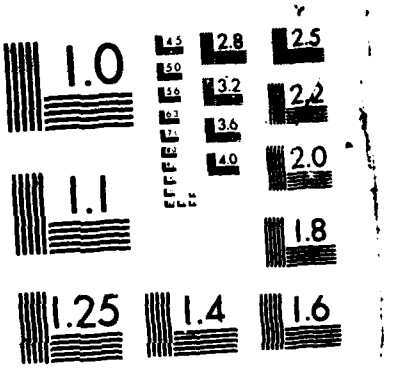
AD-A186 462 SYNTHESIS OF HIGH MOLECULAR WEIGHT PEPTIDE POLYMERS AND COPOLYMERS CONTAI (U) ALLIED-SIGNAL INC MORRISTOWN NJ BIOSCIENCES LAB J I WILLIAMS JUL 87 N00014-86-C-0484 1/1

UNCLASSIFIED F/G 6/1 NL



SECU

AD-A186 462 REPORT DOCUMENTATION PAGE

1a. (U)	1b. RESTRICTIVE MARKINGS N/A
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT DISTRIBUTION STATEMENT A Approved for public release; Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) N/A	5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A
6a. NAME OF PERFORMING ORGANIZATION Allied-Signal, Inc.	6b. OFFICE SYMBOL (If applicable) N/A
6c. ADDRESS (City, State, and ZIP Code) Biosciences Laboratory P. O. Box 1021R Morristown, NJ 07960	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research	8b. OFFICE SYMBOL (If applicable) ONR
8c. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA 22217-5000	7b. ADDRESS (City, State, and ZIP Code)
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-C-0484	
10. SOURCE OF FUNDING NUMBERS	PROGRAM ELEMENT NO 61153N; PROJECT NO RR04106; TASK NO NR441c019; WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
(U) Synthesis of High Molecular Weight Peptide Polymers and Copolymers Containing L-Dopa Residues

12. PERSONAL AUTHOR(S)
Williams, J.I., M.D. Swerdloff, I. Goldberg, A.J. Salerno and H. Bhattacharjee

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 8/86 TO 7/87		

16. SUPPLEMENTARY NOTATION
N/A

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
08			M. edulis polyphenolic proteins, L-Dopa, collagen, chemical peptide synthesis, recombinant DNA, adhesives.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The sea mussel M. edulis utilizes a polyphenolic protein in its adhesive plaque which contains tandem repeats of variants of a decapeptide sequence. We have chemically synthesized members of several families of peptides (glue peptides) related to a consensus decapeptide sequence identified by H. Waite and co-workers (Biochemistry 24, 5010, 1985) and have established methods for cleaving and purifying significant amounts of each. Chemical polymerization of model peptides using diphenylphosphorylazide activation has concurrently been studied in preparation for polymerization and lap shear adhesive strength tests for the glue peptides being synthesized. An evaluation of an alternative production method for polypeptides with repeating amino acid sequences has also begun using recombinant DNA techniques. Several expression vectors based on the regulatable lambda P_L promoter have been developed. Gene cassettes coding for analogues to the polyphenolic protein or collagen have been designed and prepared, and have been or are being cloned for expression studies. Expression studies with a cloned collagen analogue gene have revealed several potential drawbacks to high level microbial expression of genes with repeating amino acid sequences which are being addressed.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	(U)

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. E. Schmell or M. Marron or M. Haywood	202/696-4760	ONR

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Annual Report

Contract Title: Synthesis of High Molecular Weight Peptide Polymers and Copolymers Containing L-Dopa Residues
Contract Number: N00014-86-C-0484
Principal Investigator: Jon I. Williams
Biosciences Laboratory
Allied-Signal, Inc.
P. O. Box 1021R
Morristown, NJ 07960

Project Goals

To prepare novel water compatible adhesive polypeptides related to a sea mussel (Mytilus edulis) bioadhesive protein which contains L-Dopa residues and to investigate their mechanism of adhesion through determination of structure-activity relationships. This goal will be achieved by first synthesizing families of high molecular weight polypeptides or polypeptide copolymers by chemical or recombinant DNA methods. These polypeptides will contain L-Dopa residues or will have them introduced enzymatically, and will contain a significant proportion of repeating amino acid sequences typical of a collagen analogue, a M. edulis glue protein, or both. Lap shear extension adhesive testing and optical measurements of chain crosslinking will be carried out on the product polypeptides in order to better understand the function of L-Dopa and other residues in marine bioadhesives.

Summary of Project Accomplishments in the First Year

The leading edge of research effort on this project is occupied with chemical and biological synthesis of several related families of peptides, mostly decapeptides, and their subsequent polymerization and testing for adhesive properties. These peptides primarily are analogues of the consensus decapeptide sequence identified by H. Waite and co-workers [Biochemistry 24, 5010 (1985)] within the sea mussel Mytilus edulis polyphenolic protein that forms part of the adhesive plaque anchoring the mussel byssal threads to marine substrata. Polymerization of such peptides to form analogues to the polyphenolic protein is motivated by the tandem nature of repeating decapeptides within the polyphenolic protein and by knowledge that increased peptide molecular weight can be expected to enhance physical properties of these biopolymers, notably their adhesion to solid surfaces. An alternative production method for polypeptides with repeating amino acid sequences is also being evaluated which utilizes recombinant DNA techniques. The following descriptions of research results summarize progress in parallel efforts in these two areas (chemistry and biology) for the first year of this contract.

Chemistry

Six analogues of a dominant glue decapeptide isolated and sequenced by H. Waite and coworkers and one glue nonapeptide deleted for a lysine residue have been synthesized (Table I), as has a protected amino acid precursor (C, Figure 1) to several other glue decapeptides. Separate syntheses have produced a collagen analogue peptide, L-(glycyl-prolyl- proline)$_{14}$. All of the peptides

were synthesized using an Applied Biosystems model 430A peptide synthesizer employing solid phase techniques with symmetrical anhydrides of t-BOC amino acids on phenoxyacetylmethyl (PAM) derivatized polystyrene/divinylbenzene copolymer resin beads. Deprotection and cleavage of the resin bound peptides was accomplished by the use of trifluoromethanesulfonic acid (TFMSA) in trifluoroacetic acid (TFA) solutions. A policy decision was made in our laboratories before the beginning of this contract research that peptides would be obtained from PAM resins by cleavage and deprotection using TFMSA in TFA rather than anhydrous HF because of HF toxicity. The glue decapeptides (Table I) behaved surprisingly different from other peptides we have made by TFMSA deprotection and cleavage in that a large number of by-products were produced during this procedure. Careful analysis of the TFMSA cleavage products of these syntheses by fast atom bombardment MS revealed extensive t-butylation and benzylation of product peptides during cleavage. An optimized deprotection and cleavage procedure was developed that dramatically reduced this problem without sacrificing peptide yields (typical crude peptide yields are now 80-90% of theoretical). The utility of this technique was proven with peptides 4-7 and molecular weights of primary cleavage products were confirmed by coupled LC-MS and by amino acid analysis. Final purification was performed for all peptides by reversed phase, preparative liquid chromatography on a C-4 column to a purity greater than 98%.

Since solid phase synthesis of peptides cannot achieve sufficiently high molecular weights to confer desirable adhesive properties to the peptides and because the amino acid sequences of known marine bioadhesive proteins appear to contain multiple direct repeats of simple peptide sequences, we have devoted some energy to polymerizing low molecular weight synthetic peptides to significantly higher molecular weights. Successful peptide polymerization has been conducted with diphenylphosphorylazide as an activating agent using model peptides L-alanyl-glycine and L-(valyl-prolyl-glycyl-valyl-glycine). This method, which minimizes racemization and formation of urethanes, yielded products with intrinsic viscosities up to η=0.26 dL/g. and yields greater than 50% upon dialysis in buffer using tubing with a molecular weight cut-off of 8,000. Application of the Mark- Howink equation to the intrinsic viscosity value suggests the polypeptide molecular weight M_η may exceed 22,000. The polymerized polypeptide products after purification were analyzed by NMR and IR and are being studied now by gel permeation chromatography for accurate molecular weight determinations. Other chemical polymerization processes are being contemplated.

Tests for determining the lap shear strength of adhesive bonds produced by polypeptides layered between two polished aluminum plates have also begun. A control experiment with poly-L-lysine hydrobromide (MW 70,000) proved particularly interesting in that this poly(amino acid) exhibited a lap shear strength greater than 12 Kg/cm^2 on 1/64" thick aluminum plates with a 1" overlap.

Biology

A major and distinct part of our research effort is the biological synthesis in E. coli of polypeptides with repeating amino acid sequences by molecular cloning of totally synthetic genes. Advances in genetics and protein engineering have provided the opportunity to potentially design and produce a wide variety of novel polypeptides with unique physical properties.

One of our goals in this contract was to exploit recombinant DNA methods in order to produce high molecular weight peptide polymers by microbial means, focusing principally on polypeptides with tandem repeats of the collagen-like segment L-(glycyl-prolyl-proline)$_n$, sequences related to the consensus decapeptide from the M. edulis polyphenolic protein, and block copolymers of these polypeptides. Our approach has been to synthesize synthetic DNA gene cassettes with end-linked linker sequences that encode the basic repeat units of the polymers without interruption, construct appropriate expression vectors that are derivatives of the expression vector pJL6 (obtained from D. Court, NIH), clone the synthetic gene cassettes into these expression vectors and produce the peptide polymers upon induction of the synthetic gene expression system in a genetically complementary E. coli host. Another major goal has been to study the stability of these polymers and their genes in E. coli.

A progenitor expression vector from which we have derived a family of gene expression vectors was prepared from pJL6 by deleting the 1.9 kb PvuII-EcoRV fragment of pJL6 and inserting a synthetic SP6 promoter at this site followed by destroying the AvaI site and inserting a synthetic T7 promoter site at this latter position. The T7 promoter is then located upstream of the NdeI--ClaI--HindIII cloning site while the SP6 promoter is positioned downstream of the cloning site; these promoters allow rapid DNA sequencing of any foreign DNAs inserted between them. The progenitor plasmid has been designated pAV02 and has been used to construct the closely-related plasmids pASC2 and pAV2-pAV6 (Table II). In order to construct this set of plasmids, the NdeI--ClaI--HindIII region of pAV02 was deleted and replaced with specific oligonucleotides coding for the cloning sites shown in Table II. The NdeI site in all constructs contains the AUG start codon for protein synthesis while the HindIII site contains part of the in-frame UAA termination codon. Vector pASC2 had been intended for cloning of collagen analogue gene cassettes but has now been abandoned in favor of pAV2, which contains an SfiI cloning site with asymmetrical ends to ensure ligation of synthetic gene cassettes exclusively in head-to-tail orientation. The vector pAV4 contains BanII and AvaI restriction sites which will allow for the cloning of collagen analogue-polyphenolic protein analogue copolymer genes. The seven amino acid leader sequence L-(methionyl-alanyl-asparaginyl-isoleucinyl-asparaginyl-asparaginyl-arginine) in pAV5 and pAV6 has been chosen from the literature to maximize translational efficiency of any fusion protein upon insertion of foreign DNA in the appropriate vector. All cloning sites and synthetic gene cassettes have specifically been designed so as to allow the maintenance of the reading frame and amino acid sequence of adjacent gene cassettes without interruption. The structure of all expression plasmids constructed to date has been confirmed by physical mapping with restriction enzymes and DNA sequencing.

In order to easily construct genes encoding peptide block homopolymers or copolymers, we exploited the concept of gene cassettes. A gene cassette within the context of this contract is a repetitive DNA sequence of totally synthetic origin which encodes a particular peptide biopolymer and the ends of which contain DNA sequence variations which are uniquely recognized by a restriction endonuclease. The ease with which multiple numbers of such gene cassettes can be cloned in tandem to generate larger synthetic genes coding for high molecular weight polypeptides depends strongly on the choice of restriction endonucleases used in the design of gene cassettes and expression vectors. In particular, during the course of this contract we developed a

prejudice in favor of certain restriction endonucleases such as SfiI which produce asymmetric ends upon cleaving double-stranded DNA. For a collagen analogue gene cassette, therefore, we used oligonucleotides A and A' synthesized on an Applied Biosystems model 380B DNA synthesizer which form a larger repeating gene sequence coding for poly[L-(glycyl-prolyl-proline)] upon annealing and ligation. Synthetic SfiI linkers prepared from oligonucleotides a and a' were attached to the collagen analogue genes. These synthetic gene cassettes were stored frozen until needed. Also, for reasons discussed below, oligonucleotides B and B' that encode the same collagen analogue as A and A' have been synthesized and purified. Polyphenolic protein analogue gene cassettes, which we term glue cassettes, have similarly been prepared using oligonucleotides C and C' and linker DNA oligonucleotides c and c'. These glue cassettes also were stored frozen.

A: 5'-CGGGTCCGCC GGGTCCGC-3'
A': 5'-CGGACCCGGC GGACCCGG-3'

a: 5'-GGGCCGCCAG GGCCGCCG-3'
a': 5'-CGGCCCTGGC GGCCCCGG-3'

B: 5'-GGCCCACCGG GTCCGCCAGG CCCGCCGGGT CCACCGGGCC CGCCAGGTCC GCCG-3'
B': 5'-GGCCCGGGTG GCCCAGGCGC TCCGGGCGGC CCAGGTGGCC CGGGCGGTCC AGGC-3'

C: 5'-CCGACCTACA AAGCTAAGCC GTCTTACCCG-3'
C': 5'-CTTTGTAGGT CGGCGGGTAA GACGGCTTAG-3'
c: 5'-CCGACCTACA AAGCTAAGCC TAGTTACCCG-3'
c': 5'-CTTTGTAGGT CGGCGGGTAA CTAGGCTTAG-3'

The SfiI-collagen analogue gene cassettes were subsequently ligated into pAV2 and transformed into E. coli DC1138 [r^- m^+ pro leu (srlR301-recA)::Tn10 ( )] and colonies harboring plasmids containing SfiI-collagen analogue gene cassettes were identified. A number of recombinant colonies were archived following restriction mapping of their inserts for size. The largest synthetic gene cassette isolated is about 350 bp while the largest tandem arrangement of SfiI-collagen analogue gene cassettes is about 550 bp. Attempts have been made to clone the glue cassettes with MaeI linkers into the AvrII site of pAV3 without success. MaeI has some undesirable characteristics such as poor ligation efficiency that we were not aware of when we started; if further attempts fail with MaeI, we will work with an alternative linker-cloning site combination that we have designed that retains all important features of the gene cassette approach.

We have made progress in optimizing gene expression from a regulatable P_L promoter during the period when the above constructions were being made and cloned. So far, we have had moderate success in expressing a synthetic collagen analogue gene under P_L promoter control which was previously constructed in our laboratory using recA$^-$ cI857 strains where the recA mutation is either a transposon mutant (strain MH3) or a deletion (strain DC1139A). In both instances, the protein product was unstable ($t_{1/2}$ < 11 minutes). We recently completed construction of a multipurpose strain that carries a recA deletion, the cI857 mutation and an rpoH165 mutation. The rpoH165 mutation completely inhibits the E. coli heat shock system and all associated proteases. Initial experiments with this strain strongly suggest it will be helpful in expressing genes from a P_L promoter and stabilizing the induced gene products.

An alternative to constructing protease-deficient hosts is to use another gene expression system, one which is not dependent on heat induction. One alternative has been explored, the use of a chemically inducible modified lac promoter. A synthetic collagen analogue gene was moved in frame from plasmid pAC1 (cf. our original project proposal) into the commercial vector pKK233-2. A recombinant plasmid, pAC3, was identified and characterized. Preliminary expression studies with pAC3 in a $lacI^q$ (srlR-recA)306::Tn10 host show synthesis of a protein but the level of expression is still under evaluation.

Another objective of this contract has been to study the stability of gene cassettes with internally repetitive sequences. Two approaches were proposed, the sizing of cassettes by restriction analysis following long-term culture and the insertion of an antibiotic resistance gene into synthetic gene cassettes in an effort to force amplification of gene cassettes through selective pressure for increased antibiotic resistance. Collagen analogue gene cassette stability in recA hosts has proven to be easily monitored by physical mapping of gene cassettes prepared from small-scale or large-scale (i.e., CsCl- purified) plasmid preparations. Colonies harboring plasmids with larger gene cassettes (greater than about 300 bp) appear to correlate with the presence of multiple gene cassettes within these strains. Subsequent deletions within gene cassettes have also been deleted. This latter result has been observed in a variety of recA mutants, including strains entirely deleted in recA, suggesting this phenomenon is recA-independent.

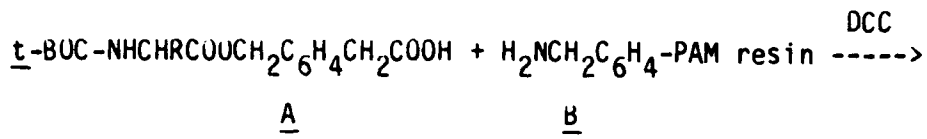
We believe the most likely explanation at this time for deletions is given by the Streisinger model [C.S.H.S.Q.B. 31, 77-84 (1966)] whereby, after DNA strand breakage or during DNA replication, slipped strand mispairing occurs at tandemly repeated sequences. The frequency of such events occurring in our gene cassettes should be inversely proportional to the complexity of the synthetic gene. The complexity of our test poly(L-glycyl-prolyl-proline) gene cassettes is extremely low (only 9 bp) and the repeat sequence contains within it the nested repeat CCGCCG. These factors may predispose StiI-collagen analogue gene cassettes to deletion events like those observed. One possible means of counteracting the tendency for deletions in our constructs is by increasing the complexity of repeating genes without changing the encoded polypeptide sequence. A compromise of optimum codon usage is required to diversify genes encoding extremely low complexity sequences such as L-(glycyl-prolyl-proline). Oligonucleotides B and B' (see above) represent an attempt by us to do this; the complexity of the product gene cassettes will be six-fold greater than those produced using oligonucleotides A and A'. We have similarly designed an improved polyphenolic protein analogue gene that is 12 times more complex than that prepared from oligonucleotides C and C'.

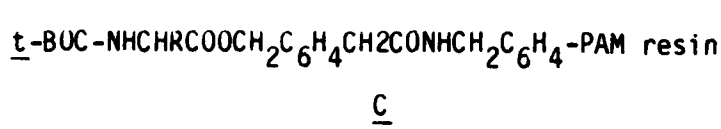
Table I. Glue Peptides for Study

1. H_2N-ALA-LYS-PRO-SER-TYR-PRO-PRO-THR-TYR-LYS-COOH
2. H_2N-ALA-LYS-PRO-SER-TYR-4-HYP-4-HYP-THR-TYR-LYS-COOH
3. H_2N-ALA-LYS-PRO-SER-TYR-4-HYP-4-HYP-THR-TYR-LYS-COOH
   /
   TFA
4. H_2N-ALA-LYS-PRO-SER-PHE-4-HYP-4-HYP-THR-TYR-LYS-COOH
5. H_2N-ALA-LYS-PRO-SER-TYR-4-HYP-4-HYP-THR-PHE-LYS-COOH
6. H_2N-ALA-PRO-SER-TYR-4-HYP-4-HYP-THR-TYR-LYS-COOH
7. H_2N-LYS-PRO-SER-TYR-4-HYP-4-HYP-THR-TYR-LYS-ALA-COOH
   /   /
   TFA   TFA

Table II. Expression Vector Constructions (Cloning Site Regions)

1. pASC2 NdeI--XmaI--ApaI--HindIII
2. pAV2 NdeI--SfiI--HindIII
3. pAV3 NdeI--AvrII(MaeI)--HindIII
4. pAV4 NdeI--BanII--AvaI--HindIII
5. pAV5 NdeI--(7 aa coding leader)--AvrII--HindIII
6. pAV6 NdeI--(7 aa coding leader)--SfiI--HindIII





[R=$(CH_2)_4NHCOCF_3$]

Figure 1

Attachment of 4-(N-t-BOC-trifluoroacetyllysine)-phenylacetic acid to PAM resin

END

DATE

FILMED

JAN

1988